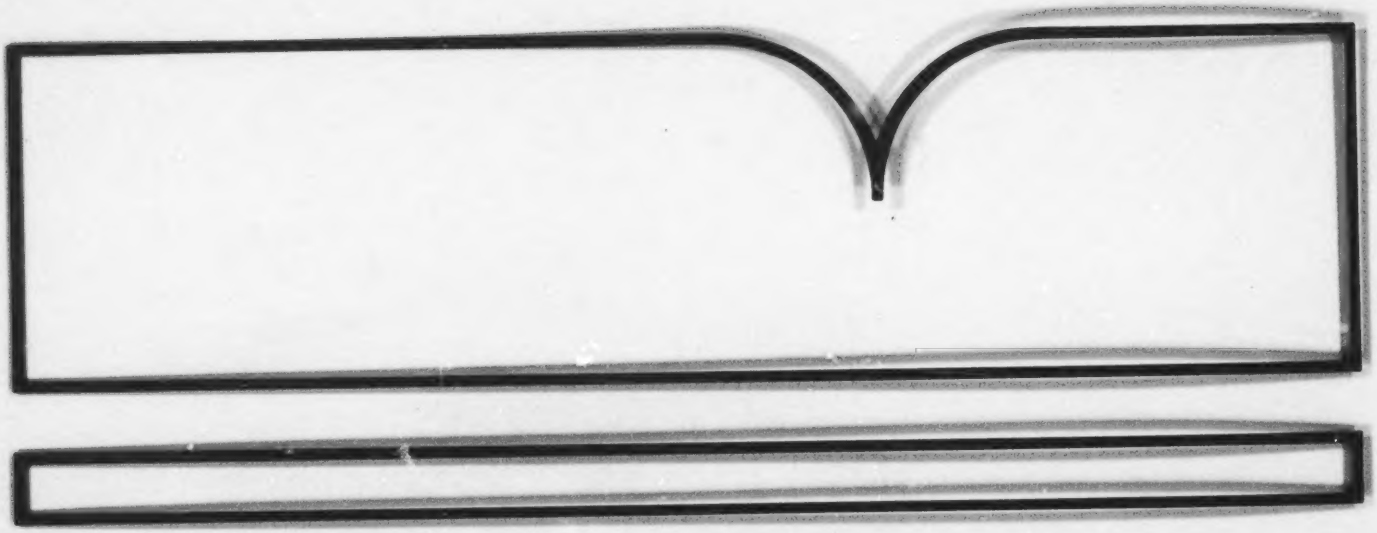
PB88-202148

Representation of Atmospheric Motion in Models of Regional-Scale Air Pollution

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

1987



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/J-87/326	2.	3. RE PB88-202148
4. TITLE AND SUBTITLE (Journal Version) THE REPRESENTATION OF ATMOSPHERIC MOTION IN MODELS OF REGIONAL-SCALE AIR POLLUTION		5. REPORT DATE 1987 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Robert G. Lamb and Saroj K. Hati*		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS same as below * Development Consultants 24-B Park Street Calcutta 700 016 India		10. PROGRAM ELEMENT NO. A 101/B/63/21 - 1335 (FY-88) 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory-RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/09

15. SUPPLEMENTARY NOTES

REF: Journal of Climate and Applied Meteorology, 26(7):837-846, July 1987

16. ABSTRACT

A method is developed for generating ensembles of wind fields for use in regional scale (1000 km) models of transport and diffusion. The underlying objective is a methodology for representing atmospheric motion in applied air pollution models that permits explicit treatment of the uncertainty inherent in the specification of atmospheric states. The nature of this uncertainty is illustrated by showing that a set of discrete meteorological observations made at a given moment in time and the diagnostic equations of fluid motion together define a manifold in function space each point of which is a possible description of the state of the atmosphere at the moment the observations were made. It is argued that hypotheses can be adduced regarding the liklihood that individual points on the manifold describe the atmospheric state at the time of the observations; but that contrary to common practice adequate information does not exist to allow one to state with certainty that a specific function is the correct description. We advance the hypothesis that the most likely descriptions of the wind field are those points on the manifold of possible states whose velocity descriptions are compatible with given kinetic energy spectra. Using methods of operations research, we develop a technique for selecting such points from the manifold and creating finite ensembles of wind field descriptions for each of N hours that meteorological data are available. By selecting one member from each of the N ensembles, we form a sequence of functions that is a possible description of the evolution of the flow field in time. We advance the hypothesis that among the set of all such sequences that one can construct, those that provide the most likely descriptions of the flow evolution are those that best satisfy the principle of momentum conservation. We show that dynamic programming is ideally suited to finding these sequences which constitute the desired ensemble of wind specifications in space and time.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	12
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE PC $12.95/MF $6.95

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

EPA/600/J-87/326
JOURNAL ARTICLE

Reprinted from JOURNAL OF CLIMATE AND APPLIED METEOROLOGY, Vol. 26, No. 7, July 1987
American Meteorological Society

The Representation of Atmospheric Motion in Models of Regional-Scale Air Pollution

ROBERT G. LAMB AND SAROJ K. HATI*

Meteorology and Assessment Division, Atmospheric Sciences Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

(Manuscript received 1 March 1986, in final form 24 January 1987)

ABSTRACT

A method is developed for generating ensembles of wind fields for use in regional scale (1000 km) models of transport and diffusion. The underlying objective is a methodology for representing atmospheric motion in applied air pollution models that permits explicit treatment of the uncertainty inherent in the specification of atmospheric states. The nature of this uncertainty is illustrated by showing that a set of discrete meteorological observations made at a given moment in time, together with the diagnostic equations of fluid motion, defines a manifold in function space, each point of which is a possible description of the state of the atmosphere at the moment the observations were made. It is argued that hypotheses can be adduced regarding the likelihood that individual points on the manifold describe the atmospheric state at the time of the observations, but that contrary to common practice, adequate information does not exist to allow one to state with certainty that a specific function is the correct description. We advance the hypothesis that the most likely descriptions of the wind field are those points on the manifold of possible states whose velocity descriptions are compatible with given kinetic energy spectra. Using methods of operations research, we develop a technique for selecting such points from the manifold and creating finite ensembles of wind field descriptions for each of N hours that meteorological data are available. By selecting one member from each of the N ensembles, we form a sequence of functions that is a possible description of the evolution of the flow field in time. We advance the hypothesis that among the set of all such sequences that one can construct, those that provide the most likely descriptions of the flow evolution are those that best satisfy the principle of momentum conservation. We show that dynamic programming is ideally suited to finding these sequences, which constitute the desired ensemble of wind specifications in space and time.

1. Introduction

The conventional approach to describing atmospheric flow in diffusion models is to represent the wind velocity at any point as the sum of two parts: a so-called mean component which is associated with the phenomenon of transport, and a random part which represents the turbulent velocity fluctuations that give rise to the phenomenon of diffusion. The mean component is treated as a deterministic function that can be derived from meteorological wind data. The turbulent wind component is considered to be a stochastic variable, whose effects on material spread must be parameterized in terms of known or measurable statistical properties of the turbulent velocity. Representing the winds in terms of a mean and random part in air pollution modeling is a technique that was borrowed from the early engineering studies of fluid motion in pipes and over flat surfaces. In small-scale flows of these types, the concept of a mean motion is well defined; if one performs moving time averages of the instantaneous velocity at a fixed point inside a pipe or in a boundary layer over a flat surface, the average velocity tends to a constant value as the averaging time is increased. (This is true, of course, only if the forces driving the flow are steady in time.) In this case, the turbulent component of the velocity is defined as the difference between the mean part and the instantaneous velocity. Averaging moments of the turbulent velocity over time yields so-called turbulence statistics which may be used, among other things, to estimate the rate of spread of material released into such flows. Early investigators of atmospheric diffusion found that these concepts provided a viable basis for estimating material dispersion over limited distances, and thus they were incorporated into early models, such as the so-called Gaussian plume formula. The Gaussian model provides estimates of time-averaged concentration of material from steady point sources in terms of three parameters: a mean wind speed $\bar{u}$, which one can estimate using wind sensors located near the source; and two parameters σ_x and σ_y, which represent the effects of the turbulent wind fluctuations on material spread. The last two parameters are treated as universal functions of $\bar{u}$, the surface heat flux, the surface roughness and the mixing depth, all of which are measurable parameters.

Models such as the Gaussian formula are valid only within distances of a few kilometers from a source. Modeling transport and diffusion over spatial scales of

* Permanent affiliation: Development Consultants, 24-B Park St, Calcutta 700 016, India.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

100 to 1000 km and periods of several days requires a fundamentally different approach; the assumptions of stationarity and homogeneity that are implicit in the definitions of the mean and turbulent wind components used in microscale models are not valid over arbitrarily large time and space domains.

Lamb (1983, 1984) studied the problem of representing the flow field in regional scale (1000 km) air pollution models and concluded that on this scale, one cannot define a deterministic component of the flow that is analogous to the so-called mean wind used in microscale models. He argued that on the regional scale, discrete meteorological data and physical principles together are capable of delineating only a *set* of flow fields, any one of which is a possible description of the flow that actually existed in the region during the period in which the wind data were collected. One can assign weights, or likelihood estimates, to each flow field in the set of possible flows and by this process arrive at a finite set or ensemble of flow fields, each of which is assumed to have a nonzero probability of occurrence. The weights one assigns to each member of the set of possible flows are based on physical principles and additional information that one might have available, such as climatological wind statistics for the region of interest.

Most existing long-range transport models use wind fields that are fit to meteorological data using so-called objective analysis methods, none of which are universally valid. In effect, objective analysis schemes assign unit probability to a single flow field in the set of possible flows, and zero probability to all others. This procedure greatly simplifies the operational task of generating wind fields for applications, but the method is not justifiable on physical grounds. The fact remains that atmospheric states cannot be specified with certainty; in applied areas, such as air pollution modeling, where the results of model simulations can have an impact on regulatory policies, uncertainties inherent in the modeling process should be treated explicitly. The objective of this paper is to develop a methodology for accomplishing this.

We want to emphasize at the outset that the focus of this paper is the development of a conceptual framework for generating ensembles of wind fields and their associated probabilities. We introduce specific techniques and procedures to illustrate basic concepts, but in their present form these procedures do not constitute an operational scheme. In a future paper, we hope to present a technique for generating wind field ensembles that can be used in routine applications of regional-scale air pollution models.

We also want to emphasize the basic philosophical difference between the method that we propose in this paper for describing atmospheric motion and the methods in current use, particularly those that use dynamical models to generate flow fields. Briefly stated, in our method, observational data play the principal role and physical laws are used to supplement the data. By contrast, in techniques that employ dynamical models, observations are either subordinated by physical principles, as is the case in so-called 4-D data assimilation models (Yamada and Kao, 1986; Stauffer et al., 1985) in which the solutions of the momentum equation are "nudged" toward the observations; or the observations are supplanted altogether, which is the case when forecasted winds are used in place of observations. Our motivation for exploring new methods arises from the belief that conventional techniques do not provide an acceptable basis for deriving the wind field descriptions needed in regional-scale atmospheric transport and chemistry models. We believe that existing methods are deficient in two basic respects. First, none of the dynamical procedures that we have mentioned is capable of producing wind fields over a period of the order of 1 week (the time interval necessary to model typical summertime air pollution episodes over 1000 km scale regions) that agree with observations closely enough to be considered credible replicas of the real atmosphere. Since observational data are the basis on which the performance of meteorological models is ultimately judged, especially in engineering oriented applications such as air pollution modeling, it stands to reason that the most desirable source of meteorological information is one that is in agreement with all available observations at all times during the simulation. In this type of application, satisfaction of physical principles is of secondary importance.

The second deficiency of conventional techniques is that all of them impose a single solution on a problem, namely, the specification of atmospheric states that inherently lacks a unique description. We mentioned this earlier when we pointed out the importance of dealing explicitly with inherent uncertainty in models that are used to develop regulatory policies.

The approach that we propose in this paper attempts to remedy both of these shortcomings of existing methods.

2. Definition of the ensemble of flow fields

For simplicity, we shall work with two-dimensional flows. In this context, we mean by a "flow field" a function pair $[u(x, y, t), v(x, y, t)]$ that describes the x–y components of the atmospheric motion at each point (x, y) in a domain D, which is generally an area of order 1000 km on a side, and at each instant t in the interval $0 \leq t \leq T$, where T may be several days or longer. By an "ensemble" of flows, we mean a set of I functions

$$[u(x, y, t), v(x, y, t)]_i, \quad i = 1 \cdots I \qquad (2.1)$$

(to be defined later) in which the ith member is different from all other $(I - 1)$ members. For each member of the ensemble (2.1), we seek a number ρ_i, such that

$$\sum_{i=1}^{I} \rho_i = 1,$$

whose value is a measure of our confidence that that particular function, $[u, v]_i$, is the correct description of the flow field in a given instance. Thinking of the ρ_is as probability values, we define the "expectation" of any property ζ of the flow fields as

$$\langle \zeta \rangle = \frac{1}{I} \sum_{i=1}^{I} \zeta_i \rho_i \tag{2.2}$$

where the angle brackets denote an ensemble average.

Our principal interest in this paper is developing descriptions of the winds for use in air pollution models. The concentrations (mass/volume) of a set of chemical species $a, b, c, \ldots$ are related to the wind fields through the mass continuity equation

$$\frac{\partial a}{\partial t} + \frac{\partial}{\partial x}(ua) + \frac{\partial}{\partial y}(va) = S_a + R_a \tag{2.3}$$

where $S_a(x, y, t)$ describes the distribution of sources of species a and R_a is a function, generally nonlinear, that describes the chemical interaction of species a with all other species. The function R links (2.3) to the corresponding equations governing the other species concentrations $b, c, \ldots$. Thus, for given emissions and reaction descriptions S and R, and given boundary conditions, there exists through (2.3) a one-to-one correspondence between members $[u, v]_i$ of the flow ensemble represented by (2.1) and members

$$[a(x, y, t), b(x, y, t), \ldots]_i, \quad i = 1, \cdots I \tag{2.4}$$

of an ensemble of concentrations, each of which is a solution of (2.3). The probability of $[a, b, \ldots]_i$ is the same as that of its counterpart $[u, v]_i$ in the flow ensemble, namely ρ_i.

The philosophy expounded in Lamb (1984) is that from an engineering perspective, the only meaningful application of models is in the prediction of properties of the concentration ensemble, such as expectations of values, because one cannot state with certainty that a particular concentration field $[a, b, \ldots]_i$ is the one that would occur for a given set of species emissions S under given meteorological conditions specified in terms of discrete observations, such as a set of wind measurements $(\hat{u}_j, \hat{v}_j)$, $j = 1, \cdots J$ made at arbitrary sites. The expectation of concentration, say $\langle a \rangle$, and its variance $\langle a^2 \rangle$, provide a best estimate of the range of values that might occur under given conditions, and consequently, they provide the best basis for the engineering decisions that must be made in the design of pollutant emissions control strategies. Our objective in this paper is to develop a working definition of the flow field ensemble (2.1) that will allow one to apply diffusion models in this role.

The flow ensemble (2.1) is defined most conveniently in terms of the discrete complex Fourier transforms of the wind components u and v, namely,

$$u(x, y, t) = \sum_{k=-M}^{M} \sum_{l=-M}^{M} U(k, l, t) e^{i(k\mu x + l\nu y)} \tag{2.5a}$$

$$v(x, y, t) = \sum_{k=-M}^{M} \sum_{l=-M}^{M} V(k, l, t) e^{i(k\mu x + l\nu y)} \tag{2.5b}$$

where M is a given integer and

$$\mu = 2\pi / L_x \tag{2.6a}$$

$$\nu = 2\pi / L_y \tag{2.6b}$$

where (L_x, L_y) are the dimensions of the simulation domain D, or larger. In (2.5) U and V are complex Fourier amplitudes. For convenience, we will drop the reference to the time dependence of these variables wherever it is unimportant in the analyses. Since u and v are real variables, we find from (2.5) that U and V must satisfy the conditions

$$U_r(k, l) = U_r(-k, -l) \tag{2.7}$$

$$U_i(k, l) = -U_i(-k, -l) \tag{2.8}$$

where

$$U(k, l) = U_r(k, l) + iU_i(k, l). \tag{2.9}$$

Similar relationships apply to V, V_r and V_i.

In order for a given mathematical function to describe atmospheric motion, it must be consistent with the equations that describe the so-called physical laws that govern fluid motion. One of these laws is the conservation of mass which, in the case of a shallow, 2-D fluid, is expressed by

$$\partial u / \partial x + \partial v / \partial y = \delta(x, y) \tag{2.10}$$

where δ is the given horizontal divergence of the flow. Taking the Fourier transform of (2.10) we get

$$ik\mu U + il\nu V = D \tag{2.11}$$

where $D(k, l)$ is the transform of the given function $\delta(x, y)$. Equation 2.11 imposes the following relationship on the transform amplitudes:

$$k\mu U_i + l\nu V_i = -D_r \tag{2.12a}$$

$$k\mu U_r + l\nu V_r = D_i. \tag{2.12b}$$

Thus in searching for mathematical descriptions of the horizontal winds, we restrict ourselves to transforms $[U(k, l), V(k, l)]$ that satisfy (2.12).

For the purposes of the present study, we will consider only nondivergent flows, i.e., $\delta = 0$, in which case (2.12) gives

$$V_r(k, l) = -\left(\frac{k\mu}{l\nu}\right) U_r(k, l), \quad l \neq 0 \tag{2.13a}$$

$$V_i(k, l) = -\left(\frac{k\mu}{l\nu}\right) U_i(k, l), \quad l \neq 0 \tag{2.13b}$$

and

$$U_r(k,0) = U_i(k,0) = 0, \quad l = 0, \quad k > 0 \quad (2.13c)$$

These conditions significantly reduce the number of independent Fourier variables that describe nondivergent, horizontal atmospheric motion.

The law of mass conservation (2.10) is but one constraint on the set of functions that describe atmospheric flow. Another constraint is compatibility with the observed wind velocities in the domain of interest. In particular, we must have

$$\sum_{k=-M}^{M}\sum_{l=-M}^{M} U(k,l,t_1)e^{i(k_\mu x_j + l\nu y_j)} = \hat{u}_j(t_1) \quad (2.14a)$$

$$j = 1, \cdots J.$$

$$\sum_{k=-M}^{M}\sum_{l=-M}^{M} V(k,l,t_1)e^{i(k_\mu x_j + l\nu y_j)} = \hat{v}_j(t_1) \quad (2.14b)$$

where $(\hat{u}_j, \hat{v}_j)$ is the wind velocity measured at point $(x, y) = (x_j, y_j)$ at time t_1; and we assume that a total of J measurements were made at this time at given sites (x_j, y_j), $j = 1, \cdots, J$. The number J and locations of the measurements may change as t changes. On expanding (2.14) and making use of conditions (2.7, 8) and its V counterpart, and constraint (2.13), we find that the flow field descriptions we seek are contained in the set of transforms (U, V) that satisfy the following system of equations:

$$\sum_{k=-M}^{M}\sum_{l=-M}^{M} [a_j(k,l)U_r(k,l) + b_j(k,l)U_i(k,l)] + U_r(0,0) = \hat{u}_j$$

$$\sum_{k=-M}^{M}\sum_{l=1}^{M} [c_j(k,l)U_r(k,l) + d_j(k,l)U_i(k,l)] + \sum_{k=1}^{M} [e_j(k)V_r(k,0) + f_j(k)V_i(k,0)] + V_r(0,0) = \hat{v}_j$$

$$j = 1, \cdots, J \quad (2.15)$$

where $a_j, b_j, \ldots, f_j$ are functions of (k, l) and the locations (x_j, y_j) of the observation stations. Note, too, that due to condition (2.8) the imaginary parts of $U(0, 0)$ and $V(0, 0)$ are identically zero. The constant terms $U_r(0, 0)$ and $V_r(0, 0)$ are given.

For a given hour t_1 the equation set (2.15) contains $2J$ equations in $4M(M + 1)$ unknowns. Recall that J is the number of meteorological stations reporting at hour t_1 and is therefore a fixed number, while M is arbitrary inasmuch as it is the total number of wave number amplitudes in the Fourier halfspace (see 2.5) that we use to represent the flow field. In principle, M should be large enough to represent all scales of motion from the regional scale down to centimer size eddies. But since in our problem the smallest resolvable fluctuation has an x-dimension of the order of

$$l_x = \frac{L_x}{M}$$

where $L_x \sim 10^3$ km, M would have to be of the order of 10^8 for each horizontal coordinate to achieve a total representation of the flow. In practice, we must limit M to a value compatible with the capacity of existing computers, say $M = 100$, and parameterize the effects of all the omitted small scales of motion by a diffusivity or perhaps a higher order closure approximation.

With M of order 100, the number of independent Fourier amplitudes in (2.15) is of order 10^4, which is two orders of magnitude larger than the number J of meteorological stations that one would expect to find in a region of 1000 km scale. In a case such as this, where the number of dependent variables exceeds the number of constraining equations, the variables are not uniquely determined. Rather, in a $4M(M + 1)$ dimensional hyperspace in which each coordinate axis is associated with one of the dependent variables, the equation set (2.15) defines a hyperplane of dimension $4M(M + 1) - 2J$ each point of which represents a *possible* description of the wind field. We say that it is a possible description because it satisfies all observations exactly and it satisfies mass conservation. However, if one were to examine the flow descriptions represented by individual points on the hyperplane, one would find that most of them are improbable flows in the sense that they contain features that one does not normally observe in the atmosphere. For example, it is possible for a flow field to contain an intense vortex between the wind-observing sites without violating either the velocities measured at the wind stations $(\hat{u}_j, \hat{v}_j)$ or mass conservation.

We add that in all our work presented here, we assume that the data $(\hat{u}, \hat{v})$ are error free. If this is not the case, the set of possible flows comprise a volume in phase space, rather than a plane, and the analyses become more complicated.

The statement that we just made—to the effect that intense vortices are not normally observed in the atmosphere—implies that we have information about the winds in addition to the set of J measurements $(\hat{u}_j, \hat{v}_j)$. Indeed, if these $2J$ data values were all we had, we would have no basis for ruling out any of the infinitely many possible flows. Additional wind information exists in the form of climatological data, aircraft measurements of velocity spectra, and the like. Our aim at this point is to utilize this additional information to diminish as much as possible the highly underdetermined nature of the flow specification. We propose to achieve this by assigning to each of the possible flows a weight ρ, which we are calling a probability, whose value is a measure of how consistent that particular flow is with additional information. The specific information that one chooses to utilize and the manner in which the estimates of ρ are associated with it are purely subjective decisions that become an intrinsic

part of the model in which the wind fields are ultimately used. In this regard, one should adopt information that is most likely to yield accurate Lagrangian properties of the flow because, as we noted earlier, the ultimate use of the wind fields is the prediction of properties of the *concentration* ensemble.

One basis for estimating the probabilities ρ_i is the assumption that lies at the heart of the so-called normal mode initialization (NMI) procedure (see Bengtsson et al., 1981 for a detailed discussion of this technique). To apply this principle to our problem, we would have to supplement the wind observations $(\hat{u}_j, \hat{v}_j), j = 1, \cdots J$ with measurements of the mass field, say $\hat{\phi}_k, k = 1, \cdots K$. Having done that and having expanded the hyperspace of velocity states to include the mass variable ϕ, we would assume under the NMI hypothesis that the only flow fields with nonzero probabilities are those that lie on the intersection of the hyperplane of possible flows and the manifold of states, termed the slow manifold by Leith (1980), that are close enough to dynamical equilibrium that they evolve slowly in time.

This hypothesis is based on the observation that atmospheric flows generally do not exhibit the sudden outbursts of gravity waves that meteorological models predict when initialized with states that are not members of the slow manifold. We add that current implementations of the NMI procedure yield a single atmospheric state rather than an ensemble of states. This is a consequence, in part, of the joint use of objective analysis schemes which, as we noted earlier, are based on assumptions that are not universally valid.

Another method of estimating the flow field probabilities, which does not require the incorporation of mass field or any other measurements, makes use of observed atmospheric kinetic energy spectra. For example, Nastrom, Gage and Jasperson (1984) found from analyses of aircraft data that in the upper troposphere both the u and v wind components have energy spectral densities that are proportional to the $-\frac{5}{3}$ power of the wave number for wave numbers $|\mathbf{k}|$ greater than about 5×10^{-5} rad m^{-1}, i.e., wavelengths smaller than about 500 km; and energy density proportional to $|\mathbf{k}|^{-3}$ for longer wave length perturbations. In the context of our flow description, these empirical spectral relationships can be expressed in the form

$$\frac{\bar{E}_{i+1}(k,l)}{\bar{E}_i(k,l)} = \left[\frac{r_{i+1}}{r_i}\right]^{-p} \tag{2.16}$$

where

$$\bar{E}_i(k,l) = \frac{1}{N_i} \sum_{k'} \sum_{l'} [|U(k',l')|^2 + |V(k',l')|^2] \tag{2.17}$$

and where the summation is over all N_i wave number vectors (k', l') lying in a ring of inner and outer radii $(r_i - \Delta r/2)$ and $(r_i + \Delta r/2)$, respectively, in wave number space. The power p can be either 3 or $\frac{5}{3}$ depending on the magnitude of the wave number amplitude r. These observations suggest the following working hypothesis:

> The probability of occurrence of the atmospheric flow field described by a given point on the hyperplane defined by (2.15) is directly proportional to the degree to which the Fourier transform associated with that point satisfies (2.16) with $p = 3$. (2.18)

In order to cast this hypothesis into an operational procedure, we must devise a method of finding those points on the hyperplane that satisfy (2.16) best. Once we have collected a sufficiently large sample of such points, say ten, we can assign each of them finite probability values which sum to 1 and assign all remaining points on the hyperplane zero probability. The finite set of probable flows then constitutes our working definition of the flow ensemble that we introduced earlier in (2.1).

Petersen (1976) and Panofsky and Dutton (1984) considered the problem of constructing an ensemble of functions that could be used to model temporal variations in wind velocity at a fixed site. They used Fourier series to represent each function in the ensemble and they adjusted both the phase and the amplitude of each harmonic to achieve statistical compatibility with empirical spectral energy relationships, such as (2.16), and with observed intermittency characteristics of turbulence. In our analysis, phase relationships among the Fourier components are constrained in part by the mass continuity equation (2.12) and in part by the wind observations (2.14). Additional constraints could be imposed to achieve other properties. In all cases, the objective is to end up with a set of functions whose properties and probability structure match those of the flow fields of interest. Since it is impossible to model all features of the flow precisely, the type and quantity of information that one uses to construct the model must be tailored to achieve the best simulation of those properties that are most critical to the particular phenomenon of interest. In the studies by Petersen and Panofsky and Dutton, the objective was to simulate the magnitude and frequency of wind gusts at a fixed location given the friction velocity. In our problem the objective is to simulate material transport given a set of discrete measurements of the wind velocity in space-time. In both cases, the models are, in essence, hypotheses, regardless of how simple or complex they might be, whose validities are ultimately judged on the basis of how accurately they describe specific atmospheric phenomena.

In our model, the actual selection of the flow fields that will make up the ensemble is accomplished using methods of operations research. This is done by defining an optimization problem whose solutions are identical to the flow fields we seek and then applying operations research methods to solve the problem. In our case, the appropriate optimization problem is the following:

Find a "design vector" $\mathbf{x}$ of dimension $4M(M+1)$ whose elements are independent Fourier amplitudes, i.e.,

$$\mathbf{x} = \begin{bmatrix} \mathbf{U}_r(k,l) \\ \cdots \\ \mathbf{U}_i(k,l) \\ \cdots \\ \mathbf{V}_r(k,0) \\ \cdots \\ \mathbf{V}_i(k,0) \end{bmatrix} \tag{2.19}$$

that minimizes the objective function

$$f(\mathbf{x}) = \sum_{i=1}^{K} |f_i(\mathbf{x})| \tag{2.20}$$

where

$$f_i(\mathbf{x}) = \frac{\bar{E}_{i+1}(\mathbf{x})}{\bar{E}_i(\mathbf{x})} - \left[\frac{r_{i+1}}{r_i}\right]^{-p} \tag{2.21}$$

and $\bar{E}_i(\mathbf{x})$ is given by (2.17) and K is an integer representing the largest number of rings in (k, l) space that one can define such that each ring contains at least an order of ten Fourier components U or V. In searching for the vector $\mathbf{x}$ that minimizes (2.20), we confine attention to those that satisfy the following constraints:

1) $[\mathbf{A}]\mathbf{x} = \mathbf{b}$ (2.22)

where $[\mathbf{A}]$ is a $4M(M+1) \times 2J$ matrix of the coefficients $a_j(k, l)$, $b_j(k, l) \ldots, f_j(k)$ in (2.15) and $\mathbf{b}$ is a vector of length $2J$ whose elements are the wind observations $\hat{u}_j$, $\hat{v}_j$;

$$2) \sum_{k=-M}^{M} \sum_{l=-M}^{M} [|U(k,l)|^2 + |V(k,l)|^2] \leq E \tag{2.23}$$

where E is a given energy value. This last constraint is added to gain control over the *total* kinetic energy of the flow fields that the optimization routine finds.

The optimization problem (2.20)–(2.23) is solved by first defining a secondary objective function

$$g(\mathbf{w}, \mathbf{x}) = \sum_{i=1}^{K+1} w_i \bar{E}_i(\mathbf{x}) = \mathbf{x}^T[\mathbf{Q}]\mathbf{x} \tag{2.24}$$

where

$$0 \leq w_i \leq 1, \quad i = 1, 2, \cdots, K+1 \tag{2.25a}$$

and

$$\sum_{i=1}^{K+1} w_i = 1. \tag{2.25b}$$

Equation 2.24 can be modified to incorporate constraint (2.22) explicitly:

$$F(\mathbf{w}, \mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T[\mathbf{Q}]\mathbf{x} + \boldsymbol{\lambda}^T\{[\mathbf{A}]\mathbf{x} - \mathbf{b}\} \tag{2.26}$$

where $\boldsymbol{\lambda}$ is a vector of length $2J$ of Lagrange multipliers.

The necessary conditions for minimization of the function F are

$$\frac{\partial F}{\partial \mathbf{x}} = 2[\mathbf{Q}]\mathbf{x} + [\mathbf{A}]^T\boldsymbol{\lambda} = 0 \tag{2.27}$$

$$\frac{\partial F}{\partial \boldsymbol{\lambda}} = [\mathbf{A}]\mathbf{x} - \mathbf{b} = 0 \tag{2.28}$$

whose solutions are

$$\boldsymbol{\lambda} = [\mathbf{B}]^{-1}\mathbf{b} \tag{2.29}$$

$$\mathbf{x} = -\frac{1}{2}[\mathbf{Q}]^{-1}[\mathbf{A}]^T[\mathbf{B}]^{-1}\mathbf{b} \tag{2.30}$$

where

$$[\mathbf{B}] = -\frac{1}{2}[\mathbf{A}][\mathbf{Q}]^{-1}[\mathbf{A}]^T. \tag{2.31}$$

The vector $\mathbf{x}$ of velocity transforms given by (2.30) minimizes the secondary objective function g defined by (2.24) while satisfying the equation set (2.15) that defines the set of possible flows. However, this vector does not necessarily satisfy either constraint (2.23) or the primary objective function (2.20), which prescribes the spectral energy density. To achieve compliance with these two conditions, we vary the vector $\mathbf{w}$ using nonlinear programming procedures until (2.20) is minimized and (2.23) is satisfied. Through this combined process, we arrive ultimately at a vector $\mathbf{x}$ that represents a single point on the hyperplane of possible flow fields and is consistent with the empirical energy spectrum (2.16). By changing the starting value of the vector $\mathbf{w}$ in (2.31) and/or the total energy E of the flow that enters in constraint (2.23), we can generate another vector $\mathbf{x}$ from the flow field hyperplane that is consistent with the energy spectrum (2.16). By repeating this process, we can produce an ensemble of flow fields $\mathbf{x}_i$, $i = 1, \cdots, I$, to each of which we can assign a probability value ρ_i depending on how well it satisfies the spectral condition (2.20), and scaled such that the sum of the probability values is unity.

It is important to keep in mind that the flow ensemble produced by the process just described pertains to a *single* hour t_1, namely, the hour that the observed wind velocities $(\hat{u}_j, \hat{v}_j)\, j = 1, \cdots, J$, were measured. Thus, our next task is to extend the procedure developed here for generating a flow ensemble at a single hour to treat time intervals of arbitrary length. This is the subject of section 3.

3. Generating an ensemble of wind fields in space-time

Virtually all air pollution model applications concerned with the design of emissions control strategies simulate meteorological conditions observed in the past. In these applications, one has available measurements of wind velocities and other meteorological variables taken at discrete points in space and time. In what is presently regarded as the more sophisticated

air pollution models, not all of the historical data are utilized. Instead, the data available for a given hour, say t_0, are used to initialize a meteorological model and the predictions of that model are used to drive the air pollution simulation during the period t_0 to $t_0 + T$, where T may be 12 h or longer. With this approach, the observations taken during the period t_0 to $t_0 + T$, and available in the data archive, are simply ignored. In effect, the physical laws and hypotheses embodied in the meteorological model *supplant* the observations. Lamb (1984) argued that in model applications of the type under discussion, the proper role of physical laws is to *supplement* observations. One way to achieve this in practice is to utilize the physical principles on which meteorological models are based to estimate the probabilities of the atmospheric states that the data and other physical laws indicate are possible states. The main objective of this paper is the development of a technique for implementing this approach.

At this point in our analysis, we have developed a technique that would allow us to generate an ensemble of flow fields at each hour for which wind data are available in a period T. That is, if T is N hours long, we can use the method developed in the previous section to create $N + 1$ ensembles

$$\mathbf{x}_i^n, \quad i = 1, \cdots I; \quad n = 0, 1, \cdots N \tag{3.1}$$

where $\mathbf{x}_i^n$ represents the Fourier transform of the ith flow description at hour t_n (see 2.19). As we indicated earlier, the number J of wind data stations and their locations (x_j, y_j) $j = 1, \cdots J$ may vary from one hour to another. We define a space-time ensemble of flow descriptions to be a set of K sequences of functions of the form

$$[\mathbf{x}_\alpha^0, \mathbf{x}_\beta^1, \mathbf{x}_\gamma^2, \ldots, \mathbf{x}_\psi^N]_k, \quad 1 \leqslant \alpha, \beta, \gamma, \cdots, \psi \leqslant I; \quad 1 \leqslant k \leqslant K, \tag{3.2}$$

composed of the members $\mathbf{x}_i^n$ of the $(N + 1)$ ensembles (3.1). Here K is an integer which, in practice, we take to be of the same order as I, say 10. Note that the upper bound on K is the total number of unique sequences of the form (3.2) that can be constructed from the set (3.1), namely,

$$K \leqslant K_{\max} = I^{N+1}.$$

With I of order 10 and N of order 24, $K_{\max}$ is a number of astronomical magnitude.

We now advance our second working hypothesis (compare with 2.18):

> The probability of a given sequence of flow fields $[x_\alpha^0, x_\beta^1, \ldots, x_\psi^N]$ is directly proportional to the degree to which that sequence satisfies the principle of momentum conservation. (3.3)

To implement this hypothesis, we will use dynamic programming and Bellman's principle of optimality (see, for example, Rao, 1979).

This procedure is best described by using an example. Consider any one of the flow states $\mathbf{x}_i^n$. According to the momentum conservation principle, this state will evolve into a new state, say $\mathbf{y}_i^{n+1}$, at time t_{n+1} where

$$\mathbf{y}_i^{n+1} = F(\mathbf{x}_i^n, \xi) \tag{3.4}$$

and ξ is a vector of specified independent parameters, such as boundary conditions, forcing fields, etc., and the function F represents the solution of the momentum equation. (For simplicity, we are assuming a 2-D isothermal fluid in which case F represents the solution of the vorticity equation.) We now define what we will call a penalty function ϵ_{ij}^n:

$$\epsilon_{ij}^n = \|\mathbf{y}_i^{n+1} - \mathbf{x}_j^{n+1}\|, \quad i \leqslant I, \quad j \leqslant I \tag{3.5}$$

where $\| \quad \|$ denotes a suitable norm, such as the sum of the squares of the vector components. In words, the penalty function ϵ_{ij}^n is a measure of the difference between the flow state that $\mathbf{x}_i^n$ would in theory become at time t_{n+1} and the jth member $\mathbf{x}_j^{n+1}$, $1 \leqslant j \leqslant I$, of our ensemble of states at t_{n+1}. Ideally, ϵ_{ij}^n should be zero for one value of j but, in general, this will never occur due to errors in the wind data and in the specifications of ξ, etc.

Let

$$\epsilon_i^{n*} = \min_j \{\epsilon_{ij}^n\}. \tag{3.6}$$

It follows that if $\mathbf{x}_i^n$ has been selected as the flow field description at time t_n the best description at time t_{n+1} is $\mathbf{x}_\alpha^{n+1}$, where α is the state for which the penalty function is minimum, namely

$$\epsilon_{i\alpha}^n = \epsilon_i^{n*}. \tag{3.7}$$

The dynamic programming procedure starts at the next to the last time step t_{N-1} and computes the minimum penalty functions for each of the I members of the subensemble $\mathbf{x}_i^{N-1}$, i.e.,

$$\epsilon_i^{(N-1)*}, \quad 1 \leqslant i \leqslant I. \tag{3.8}$$

These values tell us which state at time t_N is most consistent, from the standpoint of the momentum equation, with a given state $\mathbf{x}_i^{N-1}$ at time t_{N-1}. Once the minimum values (3.8) have been calculated, we move backward one step to time t_{N-2}. At that point, the penalty function is defined by (cf. 3.5)

$$\epsilon_{ij}^{N-2} = \|\mathbf{y}_i^{N-1} - \mathbf{x}_j^{N-1}\| + \epsilon_j^{(N-1)*}. \tag{3.9}$$

In words, the penalty function associated with a given pair of states $\mathbf{x}_i^n$ and $\mathbf{x}_j^{n+1}$ separated in time by a single time step is the *minimum* penalty that one could achieve by beginning at state i at time n and proceeding through state j at time $n + 1$, all the way to the final time level t_N. In analogy with (3.6), the minimum penalty for a given state $\mathbf{x}_i^{N-2}$ is

$$\epsilon_i^{(N-2)*} = \min_j \{\epsilon_{ij}^{N-2}\}. \tag{3.10}$$

The definitions of the penalty function (3.9) and

minimum penalty (3.10) can be applied at time step $N-3$ and to each earlier time level, in turn, until we finally arrive at values for the initial hour, namely,

$$\epsilon_{ij}{}^{0} \text{ and } \epsilon_i{}^{(0)*}. \tag{3.11}$$

Taking the minimum of the latter over all I, i.e.,

$$\min_i \{\epsilon_i{}^{(0)*}\} \tag{3.12}$$

we obtain the information necessary to define the optimal sequence of states $\mathbf{x}_i{}^n$ in space-time.

An example is illustrated in Fig. 1 for the case $I = N = 5$. Here $i = 2$ is assumed to be the state for which $\epsilon_i{}^{(0)*}$ is a minimum. Now, by definition, $\epsilon_i{}^{(0)*}$ is the minimum of all the penalty values $\epsilon_{ij}{}^{(0)}$, $1 \leq j \leq I\,(=5)$ that are associated with state i. Suppose that $j = 1$ is the state for which $\epsilon_{ij}{}^{(0)} = \epsilon_i{}^{(0)*}$. In this case, the pair of states $\mathbf{x}_2{}^0$, $\mathbf{x}_1{}^1$ are the two that are most consistent with the momentum equation. Looking next at state $i = 1$ at time t_1, we know the associated minimum penalty $\epsilon_1{}^{(1)*}$ and we know with which state j at time t_2 it is associated—$\epsilon_1{}^{(1)*} = \epsilon_{13}{}^{(1)}$ in the present example. This information tells us that the sequence $[\mathbf{x}_2{}^0, \mathbf{x}_1{}^1, \mathbf{x}_3{}^2]$ is optimal from the momentum conservation standpoint during the first three time steps. Proceeding in this way, we finally arrive at the full sequence illustrated by the connecting line segments in Fig. 1.

The beauty of dynamic programming in this particular problem is that it is able to determine the optimal sequence by evaluating only $N \times I^2$ of the possible sequences. By contrast, a straightforward approach would have to examine all $K_{\max}$ sequences. For example, in the case where $I = 10$ and $N = 24$, dynamic programming arrives at the optimal sequence after testing only 2400 of the 10^{25} possibilities! The final step in the implementation of the dynamic programming procedure is to express the state vectors $\mathbf{y}_i{}^n$, which enter in the definition of the penalty function [see (3.9)], in terms of the momentum equation.

Since the state vectors $\mathbf{x}_i{}^n$ are expressed in terms of complex Fourier amplitudes, we must express the solutions $\mathbf{y}_i{}^n$ of the momentum equation in the same form. We start with the equations for a 2-D, isothermal fluid in physical coordinates:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right) + \kappa\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) \tag{3.13a}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\left(\frac{\partial p}{\partial y}\right) + \kappa\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) \tag{3.13b}$$

where p is pressure, ρ is the fluid density and κ is the kinematic viscosity.

It is convenient to work with the vorticity

$$\Gamma = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{3.14}$$

whose governing equation is found from (3.13) and the mass continuity equation (2.10) (with $\delta = 0$) to be

$$\frac{\partial \Gamma}{\partial t} + u\frac{\partial \Gamma}{\partial x} + v\frac{\partial \Gamma}{\partial y} = \kappa\left(\frac{\partial^2 \Gamma}{\partial x^2} + \frac{\partial^2 \Gamma}{\partial y^2}\right). \tag{3.15}$$

In terms of the complex Fourier amplitudes (U, V) of the velocity (u, v), the vorticity is expressed in the form [see (2.5)]

$$\Gamma(x, y, t) = i \sum_{k=-M}^{M} \sum_{l=-M}^{M} [k\mu V(k, l, t) - l\mu U(k, l, t)]e^{ip} \tag{3.16}$$

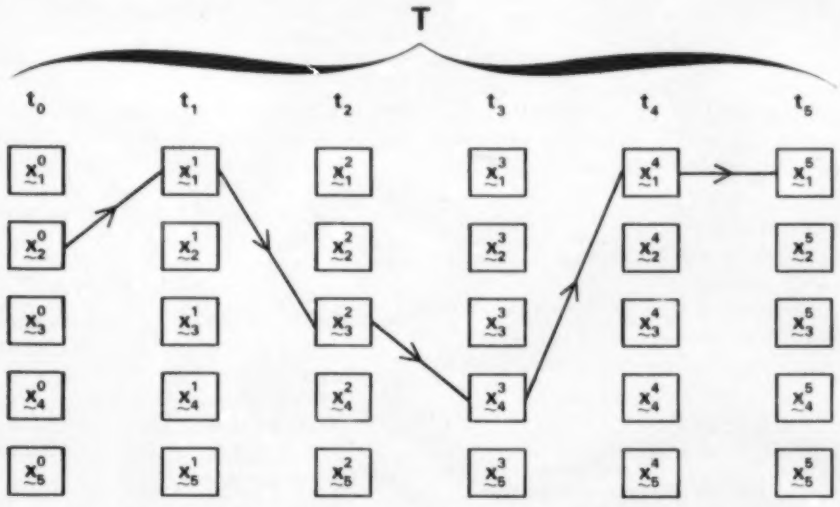


FIG. 1. Schematic representation of a sequence of flow fields that constitutes a single member of an ensemble of flows in space-time (D, T).

where

$$p = k\mu x + lvy. \tag{3.17}$$

Substituting (3.16) into (3.15) and making use of (2.13), we obtain after a considerable amount of algebra the equations governing the temporal behavior of U and V:

$$-i\frac{(k\mu)^2+(lv)^2}{lv}\frac{\partial U(k,l,t)}{\partial t} + A(k,l,t) + B(k,l,t) - i\kappa\frac{[(k\mu)^2+(lv)^2]^2}{lv}U(k,l,t) = 0, \quad l \neq 0, \tag{3.18a}$$

$$ik\mu\frac{\partial V(k,0,t)}{\partial t} + A(k,0,t) + B(k,0,t) + i\kappa(k\mu)^3 V(k,0,t) = 0, \quad l = 0. \tag{3.18b}$$

In Eq. (3.18)

$$A(k,l,t) = \sum_{k''=\max(-M,-M+k)}^{\min(M+k,M)} \sum_{l''=\max(-M,-M+l)}^{\min(M+l,M)} U(k'',l'',t) \times G_1(k-k'',l-l'',t) \tag{3.19a}$$

$$G_1(k',l',t) = -[V(k',l',t)k'\mu - U(k',l',t)l'v]k'\mu \tag{3.19b}$$

and

$$B(k,l,t) = \sum_{k''=\max(-M,-M+k)}^{\min(M+k,M)} \sum_{l''=\max(-M,-M+l)}^{\min(M+l,M)} V(k'',l'',t) \times G_2(k-k'',l-l'',t) \tag{3.20a}$$

$$G_2(k',l',t) = -[V(k',l',t)k'\mu - U(k',l',t)l'v]l'v \tag{3.20b}$$

If we initialize (3.18) with the values of the Fourier amplitudes at time t_n, i.e., if we use as initial condition (cf. 2.19)

$$\mathbf{x}^n = \begin{bmatrix} \mathbf{U}_r(k,l,t_n) \\ \vdots \\ \mathbf{V}_i(k,0,t_n) \end{bmatrix} \tag{3.21}$$

we can integrate (3.18) forward in time to obtain

$$\mathbf{y}^{n+1} = \begin{bmatrix} \mathbf{U}_r(k,l,t_{n+1})_p \\ \vdots \\ \mathbf{V}_i(k,0,t_{n+1})_p \end{bmatrix} \tag{3.22}$$

where the subscript p in (3.22) signifies that the Fourier components are predicted by the momentum equation rather than deduced from the diagnostic data. It is the predicted state vector $\mathbf{y}^{n+1}$ obtained from (3.18) for each of the states $\mathbf{x}_i^n$ that we require in (3.5) to compute the penalty functions.

The procedure that we just outlined would yield the optimum flow sequence, that is, the one with the least overall penalty ϵ. Under hypothesis (3.3), we would assign this particular flow the highest probability value ρ. By using combinations of the minimum and second-best values of the penalty function $\epsilon_{ia}{}^n$, one could generate other flow sequences to arrive finally at an ensemble of the K space-time flows, each of which is assigned a nonzero probability ρ_i, whose total is unity.

4. Conclusion

We have developed a method for deriving wind fields for use in models of regional scale (1000 km) air pollution. A fundamental difference between this and existing techniques is that our method produces ensembles of flows rather than a single flow specification. In this regard, it is consistent with the fact that physical laws and discrete meteorological data do not uniquely specify the wind field.

We demonstrated, using a 2-D fluid as an example, that a set of J discrete observations of the fluid velocity at arbitrary sites for a given instant and the equation of mass continuity together define an infinite set of functions any one of which is a possible description of the flow that existed at the time the measurements were made. We hypothesized that, of these functions, the ones most likely to describe the actual flow are those whose kinetic energy spectra conform to given empirical specifications. An optimization procedure was described that allows one to select functions possessing the desired spectral characteristics from the infinite set of possible flows. Using the sets of probable flows derived for each hour that wind data are available, we constructed sequences of flow fields to represent the evolution of velocity in time. We proposed that the probability of any given sequence of velocity fields is proportional to the degree to which that sequence satisfies the principle of momentum conservation. We showed that dynamic programming can be used to select, from the vast set of possible flow sequences, a finite subset containing the most probable of all sequences. This subset constitutes an ensemble of flow fields in space-time whose members map one-to-one onto a corresponding ensemble of species concentration distributions. Our underlying philosophy is that it is only properties of the concentration ensemble, rather than properties of any one of its members, that constitute a proper basis for decision making in applied air pollution modeling. It is only in the ensemble properties that the inherent uncertainty in specifying atmospheric flow is taken into account.

As we indicated at the outset, the techniques presented in this paper are not suitable in their present form for operational use. In a future paper we hope to develop a scheme that will allow the ideas developed here to be implemented in regional scale models of air pollution.

Acknowledgments. S. K. Hati gratefully acknowledges the support of a National Research Council-Environmental Protection Agency Research Associateship during the course of this work. Both of us are indebted

to our colleague Dr. Don Pelles for his valuable contributions to this research.

REFERENCES

Lamb, R. G., 1984: Air pollution models as descriptors of cause-effect relationships. *Atmos. Environ.*, **18,** 591–606.

——, 1983: A regional scale (1000 km) model of photochemical air pollution. Part 1: Theoretical formulation. EPA report EPA-600/3-83-035, 238 pp.

Bengtsson, L., M. Ghil and E. Kallen, Eds., 1981: *Dynamic Meteorology Data Assimilation Methods,* Springer-Verlag, 330 pp.

Leith, C., 1980: Non-linear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37,** 958–968.

Nastrom, G. D., K. S. Gage and W. H. Jasperson, 1984: Kinetic energy spectrum of large and mesoscale atmospheric processes. *Nature,* **310,** 36–38.

Panofsky, H. A., and J. A. Dutton, 1984: *Atmospheric Turbulence: Models and Methods for Engineering Applications,* Wiley and Sons, 397 pp.

Petersen, E. L., 1976: A model for simulation of atmospheric turbulence. *J. Appl. Meteor.*, **15,** 571–587.

Rao, S. S., 1979: *Optimization. Theory and Applications,* Wiley and Sons, 711 pp.

Stauffer, D. R., T. T. Warner and N. L. Seaman, 1985: A Newtonian "nudging" approach to four-dimensional data assimilation: Use of SESAME-VI data in a mesoscale model. *Preprints, Seventh Conf. on Numerical Weather Prediction,* June 17–20, Montreal, 77–82.

Yamada, T., and C. Y. J. Kao, 1986: Long-range tracer data simulations using a four-dimensional data assimilation method. Extended abstract volume. *Fifth Joint Conf. on Applications of Air Pollution Meteorology,* Chapel Hill, Amer. Meteor. Soc..

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.